DOCUMENT RESUME

ED 319 782 TM 015 009

AUTHOR Wood, Peter H.; And Others
TITLE Grading and Evaluation Practices and Policies of School Teachers.
PUB DATE Apr 90
NOTE 58p.; Paper presented at a joint session of the Annual Meetings of the National Council on Measurement in Education (Boston, MA, April 17-19, 1990) and the American Educational Research Association (Boston, MA, April 16-20, 1990).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150) -- Tests/Evaluation Instruments (160)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Elementary School Teachers; Elementary Secondary Education; Evaluation Methods; Formative Evaluation; *Grading; *Secondary School Teachers; *Student Evaluation; *Teacher Attitudes; Test Use
IDENTIFIERS Ohio

ABSTRACT

To assess teachers' grading and other evaluation practices and policies in Ohio, a 79-item questionnaire was administered to 258 elementary and secondary school teachers, and 137 other teachers were interviewed by college students taking an educational assessment class. The typical teacher assigned term grades of "A" or "B" to 56% of their students, while assigning "F" grades to 5% of their students. Grades were based primarily on student achievement that was reflected in classwork or seatwork, completion questions, homework, and matching questions, in that order. The most frequently used assessment procedures were teacher observation, in-class work, oral questions, and homework. Oral questions, seatwork, completion questions, observation of effort, and problem questions were most valued as indicators of end-term achievement and as formative indicators of progress. Most subjects preferred administrator judgments to peer ratings or student ratings of their teaching effectiveness. Student ratings were slightly more acceptable to teachers of older students. The validity of teacher judgments of achievement remains an open question. Six data tables and a 38-item list of references are included. The teacher questionnaire is provided. (Author/TJH)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED319782



U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
PETER H. WOOD
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."



# Grading and Evaluation Practices and Policies of School Teachers

Peter Wood, Thomas Bennett, Jane Wood

Bowling Green State University

Charla Bennett

Liberty Center (Ohio) Schools

A paper presented at the Annual Meeting of the National Council on Measurement in Education (Session J2); Co-sponsored by the American Educational Research Association, Division D (#27.48). Wednesday, April 18, 1990, Boston, Massachusetts.

Running head: GRADING PRACTICES OF SCHOOL TEACHERS

TM015009

BEST COPY AVAILABLE

Abstract

Two-hundred and fifty-eight Ohio school teachers completed a 79-item questionnaire, and 137 other school teachers were interviewed by college students taking an educational assessment class. The typical teacher assigned term grades of A or B to 56% of their students while assigning F grades to 5% of their students. Grades were based most upon student achievement that was reflected in classwork or seatwork, completion questions, homework and matching questions--in that order. The most frequently employed assessment procedures were: teacher observation, in-class work, oral questions, and homework.

Oral questions, seatwork, completion questions, observation of effort, and problem questions were most valued as indicators of end-term achievement--and also as formative indicators of progress.

A majority preferred administrator judgments to peer ratings or student ratings of their teaching effectiveness. Student ratings were slightly more acceptable to teachers of older students.

The validity of teacher judgements of achievement remains an open question.

## Grading and Evaluation Practices and Policies of School Teachers

Assessment and evaluation tend to be an integral part of most school-based instruction. Hopkins, Stanley and Hopkins (1990, p. 7) state that, "The process of education depends extensively on assessment and measurement and on a related procedure--evaluation." Classroom evaluation has been defined by Gronlund and Linn (1990) as "the systematic process of collecting, analyzing, and interpreting information to determine the extent to which pupils are achieving instructional objectives" (p. 5). At the end of each academic term, the summary results of a multitude of evaluative activities are reported to students and to others in what Gronlund and Linn have described as "one of the more frustrating aspects of teaching.... There are so many factors to consider...and so many decisions to be made" (p. 427).

### The Major Purposes of This Study

A major purpose of this study was to gain some understanding of how a convenience sample of experienced school teachers distributed their end-term grades across the traditional A-B-C-D-F grading scale. A second purpose was to determine how different teachers reported weighting student achievement, effort, and potential in their grading decisions. Additional foci were upon: (1) the frequency with which teachers used common measures for summative--and formative--purposes; (2) teachers' attitudes toward

a range of assessment issues; and teachers' choices of ways that they would most prefer to have their own teaching evaluated. Because two of the authors teach about testing to pre-service teachers, the study could be categorized as "action research," or "teacher research" (Cross & Angelo, 1988; Cochran-Smith & Lytle, 1990).

## Grading Practices

The types of information put on report cards appear to have been the focus of two thorough and relatively recent surveys. Gronlund (1985, p. 441) cites a 1970 study by the NEA Research Division that indicated that 72% of elementary teachers and 83% of secondary school teachers used letter grades (e.g., A, B, C, D, F) on report cards. Parent conferences (60% elementary; 10% secondary), written descriptions of performance (24%, 10%) and number grades (10%, 9%) replaced or supplemented the letter grades for some teachers. Kunder and Porwoll (1977) conducted a mailed survey of a stratified random sample reflecting the 11,509 American school systems that had over 300 pupils. Their usable responses (32%) presented a picture in which letter grades were used by: 90% of senior high schools, 91% of junior high schools, 69% of upper elementary grades, 37% of primary grades, and 8% of kindergarten teachers. Parent-teacher conferencing was the most frequently employed reporting system in the K-6 grades, with checklists and rating scales the next most frequently used system.

Although the types of information provided to parents is relatively well established, the sources and distributions of this information have not been as clearly identified. A very experienced observer of educational efforts, Paul Dressel (cited in Milton, Pollio & Eison, 1986, p. 23) has described the course grade as "An inadequate report of an inaccurate judgment by a biased and variable judge of the extent to which a student has attained an undefined level of mastery of an unknown proportion of an indefinite material." Stiggins, Frisbie and Griswold (1989) reported that "a review of grading research literature yields little that is of value in answering" grade-related questions (p. 6). They reported their opinion that neither "a summary of actual practices teachers use to generate grades" (p. 6) nor "an analysis of the underlying assumptions and philosophies teachers use in the grading process" (p. 6) is available in the research literature. Most of the 15 high school teachers involved in their case study employed paper-and-pencil tests as a primary contributor to grades, but two also included results of oral questions, nine may have included daily written assignments, and almost half used a range of other performance measures. Although half used fixed percentage grade cut-offs, different teachers often used different cut-off scores.

Gronlund (1985); Gronlund and Linn (1990); Hopkins, Stanley and Hopkins (1990); Popham (1990); and Stiggins, Frisbie and

Griswold (1989) all support basing the grades exclusively upon demonstrated student achievement. However, Bowling Green State University faculty who teach future teachers of the handicapped have strongly suggested that students of different abilities should be provided with different assignments, different types of placement, and different grading systems. Both Ohio law (amended House Bill 455) and Federal law (PL 94-142) require the use of student-specific goals, objectives, assessment practices and instructional procedures for those handicapped students who may total approximately 11% of the school student population (see Biehler & Snowman, 1986, p. 188). It would seem to be difficult to apply the same absolute or normative grading standards to students who are provided with different objectives and with different teaching procedures. The reporting of pupil "progress toward objectives" should differ for teachers who have the same objectives for all students and for teachers who establish different objectives for different students. This difference in approaches to grading should be especially obvious in a comparison of teachers of students who vary most in initial abilities (e.g., handicapped students, athletes, musicians)--and teachers of more homogeneously grouped students in traditional academic subjects.

Two additional factors can further complicate the grading process: (1) Some music and physical education teachers are responsible for grading several hundred students of vastly

differing initial abilities. The large numbers of students and the little time spent with each may cause a shift toward grading based upon attendance or apparent effort. Class size may also influence some special education teachers (who teach a small number of extensively measured students). For them, grading on improvement may be appropriate. (2) Grading procedures also may differ across fields that differ in the types of students that can succeed in the field. Milton, Pollio and Eison (1986, chapter 2) argue that certain academic disciplines (notably the physical sciences and mathematics) and certain gate-keeper courses (e.g., organic chemistry for pre-medical students and statistics for psychology students) are noted for rigorous, differential grading standards. Goldman and Slaughter (1976) presented the results of a series of studies of college grades to support Goldman's thesis that fewer high grades are assigned in those fields that draw the most able students. If this also were true of pre-college grading systems, one could expect higher grades from teachers who teach handicapped students or very young students or from teachers who teach performance-related subjects (band, chorus, physical education) in which most students do not function at the levels expected from the highly talented or skilled students who are coached by the same teachers in co-curricular activities. In each case, the obviously lesser abilities of the students could lead to

grading systems that reward effort and improvement--and that tend to produce higher distributions of A or B grades.

Evaluation Techniques: Frequency of Use

Ron Marso and Fred Pigge have reported the results of several surveys of Ohio school teachers. In 1985, Marso reported that the typical teacher (n = 123) administered over 99 tests during the school year. The most common item type was the completion or short-answer question--followed by problem questions, matching, multiple-choice, essay, and true-false questions. Matching questions were rated as the most useful, followed by multiple-choice, completion, true-false, problems, and essays, in that order. Almost all (96%) used some self-constructed items, and 97% used test questions as a part of assigning grades. Question use and preference varied across grade levels. Marso and Pigge (1987, 1988a, 1988b) also reported the results of a survey of another 326 Ohio K-12 teachers. They found that the "average" teachers: (1) administered 51 formal tests each year; (2) wrote most of their own test questions; (3) most frequently used multiple-choice, matching, and short-response items; (4) infrequently employed longer essay items; (5) infrequently calculated means, standard deviations, reliability coefficients, or item analyses; and (6) varied in their practices across grade levels. As has been found by almost all previous studies (Fleming & Chambers, 1983; Gullickson & Ellwein, 1985), Marso and Pigge (1988a, 1988b)

examined a wide range of samples of teacher tests and discovered that the great majority of the questions appeared to reflect relatively low levels of cognitive skill.

Kirby and Oescher (1989) asked 16 high school science and 19 high school math teachers to complete a questionnaire and also to provide their most recent unit or term test. The teachers reported writing about 65% of the 1400+ questions on their tests. Short-answer questions accounted for more than 60% of the items with multiple-choice (20%), matching (15%), true-false (8%), and essay (only 4 of 1400+ items) questions also represented. The teachers tended to overestimate their use of essay questions and underestimate their use of short-answer items--possibly because of different definitions of essay. Less than 7% of the questions were judged to assess cognitive levels beyond the comprehension level. The math and science teachers differed as to question level with the majority (78%) of the math questions written at the comprehension level while 78% of the science questions were written at the knowledge level. Over 70% of all tests were employed primarily for assigning grades rather than for more formative reasons.

In summation: (1) There are several studies that focus upon the types of procedures teachers use to evaluate their students. (2) These studies appear to indicate that teachers of different levels and different subjects seem to employ somewhat different

types of test questions. (3) There are fewer studies that relate testing procedures to summative evaluation and grading--and even fewer relating testing to formative procedures. (4) There are logical and situational reasons for different teachers to distribute grades differently--and to weight different factors in their grading. (5) Teachers create many tests each year, employ mostly objectively scoreable items, and use the results more for summative/grading purpose than for formative/teaching purposes.

Evaluation of the Teacher by Others

Certification requirements for Ohio teachers specify that teacher-evaluation procedures be included in the pre-service curriculum. A survey by Wood, Stang and Wood (1980) reported that the majority of the faculty of one midwestern college preferred to be evaluated by their students--if they had to be evaluated at all. The majority of the evaluations of elementary and secondary teachers is conducted by school administrators although some of the newer Ohio induction programs (e.g., The Toledo "Intern/Intervention/Evaluation" Plan, 1985) employ experienced, peer teachers as the major evaluators of first year teachers.

Specific Questions

The following questions were asked of Ohio teachers in order to gain some understanding of their grading, measurement and evaluation practices.

1. What percent of their students would these teachers assign to A, B, C, D, and F grade categories--if they were required to assign letter grades?
2. What student characteristics (achievement, motivation, attendance, completion of assigned work, improvement, class participation) would contribute to student assignment to a particular grade category--and how would these factors be weighted?
3. How frequently are various types of measurement devices employed by these teachers?
4. Which devices are preferred for use in summative evaluation--and for formative evaluation?
5. What are the teachers' attitudes toward various evaluation issues--including the use of teacher-made, district-wide and state-wide tests?
6. How would teachers prefer to be evaluated?
7. Do teachers teaching different types of classes or grade levels differ in their answers to the above questions.

Hypotheses

It was hypothesized that: (1) teachers of handicapped students and teachers of performance-based classes (e.g., music and physical education) would assign higher grades to more students; (2) these two groups of teachers would weigh improvement more heavily in their term grades; (3) elementary teachers would

weigh homework and classwork more than secondary teachers; and (4) objectively scored questions would be employed for end-term evaluation and application to report cards while a wider range of less formal and more subjective procedures would be employed for formative evaluation.

## Method

### Subjects

Two hundred and fifty-eight school teachers completed an 79-item survey. Approximately 50% taught elementary students (K-6), and 50% taught secondary students. Approximately 12% worked primarily with handicapped students, 6% taught vocational courses, and 5% taught various performance-related subjects (music, art, physical education). The majority (77%) taught regular students in academically-oriented classes. Table 1 presents much of the biographic data describing the teachers responding to the questionnaire.

An additional 137 school teachers were interviewed by college students who planned to become school teachers. Forty-one percent taught in grades K through 6, and 59% taught older students. Only 5% taught handicapped students. Descriptions of these teachers and their views will be presented in a subsequent paper.

- - - - - - - - - - - - - - -

Insert Table 1 about here

- - - - - - - - - - - - - - -

## Measures

The two data-gathering instruments were an 79-item, four-page questionnaire and a one-page interview guide.

The Questionnaire. The questionnaire consisted of: nine general biographic questions; two questions asking teachers to specify a subject and grade level that the teacher is best able to describe; five questions asking teachers to describe the percentage of students that they would assign to five letter grades (A through F); an eight-item question asking for sources of the grade-assignment decision; a list of 21 typical types of measurement devices (formal and less formal) with a request for a frequency-of-use indicator; six questions related to formative and summative assessment; 17 attitude questions; and an 11-part segment asking teachers how they would prefer to be evaluated. Most of the items had undergone trials with groups of college students or faculty. Questions concerning teacher goals were eliminated from the final questionnaire because trial groups of college students and pre-college teachers did not appear able to choose among the goals provided on the trial questionnaire.

The Interviews. Undergraduate juniors and seniors in pre-service teacher education programs were each asked to interview two exemplary teachers of subjects and grade levels similar to those the students planned to teach. The names and addresses of the interviewed teachers were included in the

interview in order to reduce the use imaginary interviewees. The results of the interviews were presented in oral and in written form in one of the author's classes (over a period of two years). Those interviews that were complete and that could be described in detail by the interviewer will be described in a subsequent paper.

## Procedures

The authors asked colleagues to identify experienced teachers in non-urban, northwest Ohio who were believed to be successful and competent teachers. An attempt was made to oversample special education and music teachers. Two hundred and fifty-eight (of 600) completed the questionnaire. Answers to those questions which involved summing percentages to 100% were recalculated so that they really did sum to 100%.

Pre-service teacher education students were asked to interview their best two former teachers--ideally teachers of levels and subjects that the students hoped to teach in the near future. The most coherent and complete 137 interviews were sorted into grade levels and types of courses taught. An initial content analysis of the interviews is still in progress.

## Results

Table 1 presents much of the biographic detail of the surveyed teachers. The mean teaching experience of these teachers was 12.8 years.

Results from the Survey: Allocation of A, B, C, D, and F Grades

Table 2 presents the mean percentages of students that were allocated to the five grade levels (A, B, C, D, F) by different groups of teachers. The typical (median) teachers assigned A, B, C, D, and F grades to 20%, 30%, 30%, 10%, and 5% of their students. Somewhat over half (55%) of their students received either an A or B grade while around 13% to 15% received grades in the D–F range. However, there was a notable difference across individual teachers in their grade assignments. One teacher assigned As or Bs to only 4% of his or her students, three other teachers assigned As or Bs to all students, and 25% assigned As or Bs to a great majority of their students (over 73% of them). Four teachers assigned no A grades while 10% of the teachers assigned grades of A to half or more of their students. No one assigned A grades to all students, but two did report assigning A grades to 90% of their students. Elementary (K-6) teachers tended to assign more A grades (mean = 27%, sd = 19%) than did secondary teachers (mean = 22%, sd = 19%). The difference was modest but statistically significant ($F = 4.2$, $p < .05$, $df = 1, 241$). Similar differences were found in the assigning of B grades (elementary mean = 34%, secondary mean = 26%, $F = 16$, $p < .0002$). Secondary teachers assigned about twice as many D grades (13% vs 7%, $F = 37$) and F grades (7% vs 3%, $F = 26$) as did their elementary (K-6) colleagues. The assignment of failure grades

jumped in the seventh (5.4%), eighth (7%), and ninth grades (12%) from a typical 1% to 3% rate in the earlier grades--only to fall back to 6% in the last three years of high school.

Different Types of Teachers. "Performance" teachers (music, art, physical education) tended to assign more As and Bs (70%) than teachers of more traditionally academic classes (54%). Special education and vocational teachers tended to fall in between "regular" and "performance" teachers in their grading patterns. Table 2 presents the grading distributions of the 242 teachers who answered the grading questions (16 reported that they did not assign grades and couldn't even imagine what patterns they would produce if asked to assign grades). Table 2 also presents the mean grade assignments of the four types of teachers--and of teachers teaching four class levels of students (K-3, 4-6, 7-9, 10-12).

- - - - - - - - - - - - - -

Insert Table 2 about here

- - - - - - - - - - - - - -

## Factors that Contribute to Grades

"The student's demonstrated achievement of the course's academic objectives (test scores, project grades, daily work, etc.)" was the major factor reported to contribute to the grades of the great majority of the teachers surveyed. Sixty-five percent of the typical (median) teacher's grade was reported to

have been based on this single factor. "Completion of assigned homework and classwork" was the second most heavily weighted factor (median = 10%, mean = 17%). Motivation and effort (3% to 5% weighting), "improvement or working up to potential" (1% to 5%), and class "participation/behavior" (2% to 5%) were weighted to a considerably lesser degree by the typical (median, mean) teacher. Attitudes toward class and school, attendance, and completion of extra credit assignments were reported to carry little weight in the typical teachers' grading process. Elementary teachers reported weighting classwork and homework (mean = 12.1%, sd = 13.6%, $n$ = 102) only about half as heavily as secondary teachers (mean = 22.2%, sd = 18.9%, $n$ = 101, $F$ = 19, $p < .001$).

Teachers of performance classes ($n$ = 11) tended to weight achievement of academic objectives only half as heavily (mean = 23%) as other teachers. Participation and behavior in class (18%), student attitudes (12%), and improvement (9%) were more heavily weighted by performance teachers than by most other teachers. Improvement or working up to potential (11% mean weighting vs 5% to 6% for other teachers) also was weighted more heavily by teachers of handicapped students. Table 3 presents the mean weights assigned to these eight factors by different groups of teachers.

- - - - - - - - - - - - - - -

Insert Table 3 about here

- - - - - - - - - - - - - - -

Frequency of Use of Measurement Techniques

Table 4 presents the reported frequency-of-use of 20 types of measurement techniques. The majority of the teachers reported using observation of effort (67%), in-class work (63%), and oral questions (60%) on a daily basis. Approximately 80% reported using seatwork (88%), observation of effort/motivation (80%), and homework (78%) on a daily or weekly basis. Half to two-thirds reported using oral questions (68%), fill-in or completion questions (57%), and problem questions (50%) on a daily or weekly basis. Various types of personal-social measures (sociograms, guess-who questions, attitude measures, anecdotal records, and interest inventories) were not employed at any time by most teachers. The least used measure, however, was the "longer-essay question" that 72% reported would probably not be used in the class and level with which the teacher was most familiar. Elementary teachers (K-6) tended to employ "guess-who questions," anecdotal records, sociograms, seatwork, problem questions and oral questions significantly more ($p < .01$) than secondary teachers. Secondary teachers (7-12) tended to use essays more. Vocational and performance teachers tended to use homework less frequently and performances more.

- - - - - - - - - - - - - - -

Insert Table 4 about here

- - - - - - - - - - - - - - -

Uses of Measurement Techniques

Summative Evaluation. Teachers were asked "Which of the above approaches (the 20 measurement techniques) tends to give you the best idea of how well objectives in your class have been met by the end of each grading period (choose one, two, or three techniques)?" Forty-one percent chose "oral questions;" 38% chose "in-class or seatwork;" 28% chose "fill-in or completion questions;" 22% chose "observation of student effort and motivation;" and 21% chose "problem questions." Homework assignments, matching questions and short essay questions were chosen by from 15% to 17%.

Teachers also were asked a parallel question, "Which procedure(s) are most helpful when you have to make out report cards?" Seatwork (40%), fill-in/completion questions (31%), homework (26%), and matching questions (18%) were reported most frequently. Observation of effort (15%) and multiple-choice questions (14%) were reported by many.

Formative Evaluation. Which procedure is best for telling you, the teacher, if students are learning? Responses to this question indicated that 39% preferred oral questions, 21% used in-class or seatwork, and 15% used fill-in/completion and/or

problem questions. Twelve percent selected observation of student effort and/or short essay questions. Student learning was reported to be most helped by: oral questions (31%); in-class or seatwork (25%); projects (16%); and fill-in and/or problem questions and/or homework (11% to 12% each).

Teacher Attitudes toward Testing and Evaluation

Table 5 presents teacher agreement with 17 statements concerning testing and evaluation.

- - - - - - - - - - - - - - -

Insert Table 5 about here

- - - - - - - - - - - - - - -

Agreement: Over 90% of the teachers agreed that report card grades should reflect how much students have learned--and that completion of homework should be included in subject grades. Eighty-five percent agreed that grades on report cards should reflect effort as well as levels of achievement. Approximately 70% reported that: (1) they would include more projects (if they had time for grading); (2) teachers who teach similar classes in their schools tend to use similar grading systems; and (3) teachers should set different objectives for different students--and grade according to how these objectives are met. Slightly over 60% would also ask students to do more writing on exams (if they had more time). Fifty-two percent tend to set different objectives for students with differing abilities.

Approximately half reported that they tended to grade less able students more leniently and have different assignments for students with differing abilities.

Disagreement: Approximately 80% of the 205 teachers who responded to the attitude statements disagreed: (1) that they were uncomfortable with their grading procedures; (2) that tests tend to do more harm than good; and (3) that, if possible, they would not give any tests at all. Approximately 70% did not believe that state-wide or district-wide testing would improve education. Questions about these two procedures also drew high responses of "I really can't answer this questions" (23% and 24%). Approximately half of the teachers disagreed that: (1) report card grades tend to be more accurate when based mostly on test or quiz scores; (2) attendance should be included in subject grades; (3) they tend to have different assignments for students of different ability; and (4) they tend to grade less able students more leniently.

In summation, although teachers were comfortable with their grading procedures and believed grades should reflect how much students have learned, they also agreed that grades should reflect homework and effort. Regarding tests, they were pro-test but did not necessarily believe that grades were more valid when based mostly on test scores. Few believed that standardized tests (district or state) would improve education. Using different

grading systems for different students was viewed differently by different teachers.

Evaluating the Teacher

Teachers were provided with some of the sources of information that could be employed in an evaluation of their own teaching, and they were asked to weigh the different types or sources of information. The mean weights of these choices are presented in Table 6.

- - - - - - - - - - - - - - -

Insert Table 6 about here

- - - - - - - - - - - - - - -

The preferred approach to teacher evaluation was "class visits and ratings by the principal or other administrator" with a mean weight of 42%. Only 4% of the teachers failed to include this source of evaluation of themselves as teachers. Compliance with administrative regulations was included by all but 20%. Most teachers included teacher visits and ratings (61% included it) and student ratings (58% included it). However, these two sources were weighted very lightly. Two approaches commonly associated with increases in teacher salaries--years of experience and additional education--were not weighted at all by nearly 60% of the teachers. The weighting assigned to student evaluation tended to be somewhat positively correlated with grade level taught ($r = .19$, $df = 184$, $p < .005$) but not associated with years of

experience or the percentage of A and B grades assigned by the teacher.

### Interview Results

The volume of information provided by the interviews is still undergoing analysis and its results will be available at a later date.

## Conclusions and Discussion

### Grade Distributions

Individual teachers and groups of teachers differed considerably in their assignment of course "letter" grades. The average teacher produced a somewhat more lenient version of the "relative" grade distribution suggested by Gronlund and Linn (1990, pp. 441-442) for "an introductory course." However, individual teachers varied widely in their suggested grade distributions.

As was hypothesized, the 13 performance teachers (42% As; 70% As or Bs) tended to be considerably more liberal in their grading policies than the 191 less specialized teachers (22% As). The vocational teachers also proved more lenient in their grade assignments while the special education teachers proved to be much more similar to the non-specialized teachers in their grade distributions. As one of the more liberal graders reported, "The only students who fail my classes are the ones who do not make it to class." The primary grade teachers were less likely to employ

letter grades (Kunder & Porwoll, 1977) and tended to be more generous when asked to report the grades they would give if required to do so (69% As or Bs; 8% Ds or Fs). Junior high school teachers (46% As or Bs; 24% Ds or Fs) tended to be the least generous. What effects these lower grades have upon drop-out percentages is not clear. In spite of this diversity, when asked if colleagues who teach similar classes tend to use similar grading systems, 23% didn't know and 68% agreed that they did.

Although some schools, perhaps a majority, may have clearly understood and agreed-upon grade assignment systems, the comment volunteered by one of the teachers may better describe many schools. This teacher wrote: "I teach with six other teachers. We all have different ways of grading." Milton, Pollio and Eison's (1986) beliefs about the impact of context upon grades was clearly echoed by the teacher who wrote: "After teaching the last seven years in three different states and five different school districts--and prior to that in two different districts in one state--my conclusion is that my 'expectations' differ as widely as the caliber of the students that I teach.... I come into my classroom with no preconceived percentage of students getting any one grade.... Grades relate closely to the level I taught and the socioeconomic level of the district I was in." Apparently even some individual teachers may grade differently from class to class and from year to year. Much of the diversity in grade assignments

may be attributable to the fact that our respondents were drawn from many different districts--and from different grade levels and teaching areas. The reported similarity of grading systems within schools or groups may refer to the common school practice of translating particular ranges of percentage scores into letter grades (e.g., 93% to 100% = A).

Factors that Relate to Letter Grades

The great majority of the teachers reported that "the students' demonstrated achievement of the course's academic objectives" was the dominant (60+%) factor used to determine letter grades. Completion of assigned homework and classwork was assigned somewhat less than half of the remaining weight. As was hypothesized, student improvement was more than twice as important to special educators and performance teachers (11% and 9%) than to other teachers (4%). Although it was not hypothesized, it was found that participation and behavior in class was much more important to the performance teachers (18+%) than to the other special or specialized teachers (6%)--or to the less specialized teachers (4%). Since behavior is more obvious to these teachers, this finding appears logical. For unknown reasons, secondary (7-12) teachers weighed homework and classwork nearly twice as heavily as elementary teachers did when assigning grades.

Use of Different Evaluation Techniques

As could be expected, the less time-consuming and immediate-result measurement techniques (teacher observation of effort and motivation; evaluation of class and seatwork; and use of oral questions) were used on a daily basis by more than 60% of the teachers. Evaluation of homework was used by nearly half (45%) of the teachers on a daily basis.

Test questions that required students to supply answers (fill-in/completion, and "problem questions") appeared to be the next-most-frequently employed measurement approach--usually used on a daily or weekly basis by half or more of the teachers. Projects dominated the more formal, monthly-use category with 75% reporting daily, weekly or monthly use. Multiple-choice questions (72% at least monthly use), fill-in/completion questions (74% use), matching questions (64% use), and problem questions (63% use) joined oral questions (72%) homework (83%), and seatwork (89%) as approaches used by the greatest percentage of teachers at least once each month. Anecdotal records (27% use on a monthly basis) and self-esteem measures (27% monthly use) were the most used of the personal-social evaluation approaches. As was reported by Kirby and Oescher (1989), the use of longer essay questions was relatively rare (72% reported no use)--falling into a non-use category shared by the "guess-who" technique (65% non-use) and the sociogram (70% non-use). While projects were

reported to be used by 75% of the teachers on at least a monthly basis, papers (reports, themes, and term papers) were only used by 30% on a monthly basis--by about the same percentage who reported using self-esteem measures and anecdotal records (27%). It is unclear as to exactly what kinds of self-esteem measures are used with such frequency--or why projects are favored over papers.

Formative and Summative Evaluation

It was hypothesized that the less formal approaches would be employed for formative evaluation ("are students learning") while the more formal approaches would be used "at the end of each grading period" and "when you have to make out report cards." As one teacher reported, "I find it difficult to justify my grades without actual paper work to back up my cards.... I would like to feel more comfortable with using my opinions and observations to grade my students.... However, in this day of aggressive parents and having to prove everything, I do not feel that I could include this (opinion and observation) as a part of my grades." However, when teachers were asked to name those approaches that give them the best idea of how well objectives were met by the end of each grading period, oral questions and in-class work or seatwork were the most common nominees--followed by fill-in/completion questions and problem questions. When report cards were to be completed, in-class or seatwork was nominated most frequently, and with over twice the frequency of anything other than (in order of

nomination) fill-in/completion questions, homework assignments, and oral questions. Formative evaluation functions were most commonly accomplished through use of oral questions and class/seatwork. As was reported by Stiggins, Frisbie and Griswold (1989), these teachers do not seem to differentiate between formative and summative evaluation--while also seeming to prefer using the immediate and often subjective procedures to more formal testing procedures for both functions.

Teacher Attitudes

Two potentially contradictory statements drew over 90% agreement: (1) Grades on report cards should reflect how much students have learned; and (2) Completion of homework should be included in subject grades. Quite apparently, homework is considered to reflect "how much students have learned."

Another conflict appeared in that eighty-five percent of the teachers also agreed that grades on report cards should reflect effort as well as levels of achievement. Although grades should reflect achievement, and achievement was weighted more than three times as heavily as any other factor, effort was also considered--even if not considered too much (5%).

Approximately 70% agreed that: (1) they would include more projects in their class if they had more time; and (2) that teachers should set different objectives for different students and grade according to how well these different objectives are

met. The apparent value for non-test performance and for considering student differences also may underlie these positions.

Two statements produced the strongest disagreement: (1) "Tests tend to do more harm that good" (82% disagreement); and (2) "If possible, I would not give any tests at all" (78% disagreement). Two other statements produced a combination of disagreement and uncertainty from teachers in a state that is on the verge of mandatory, high-stakes, standardized testing: (1) District-wide testing tends to improve education (67% disagreement and 23% "really can't answer"); and (2) State-wide testing tends to improve education (68% disagreement and 24% uncertainty). As has been found by Gullickson (1984) and by Hall, Carroll and Comer (1988), these teachers apparently value tests as potentially useful devices, but tend to distrust the results of tests, especially those made by "outsiders."

Evaluation of the Teacher

More teachers volunteered comments concerning the evaluation of teachers than concerning their evaluation of the students in their charge. Like the college teachers described by Wood, Stang, and Wood (1980), these teachers found many problems to be associated with teacher evaluation. Unlike the college teachers who tended to prefer students as judges of their effectiveness, more of these teachers of younger students nominated administrator visitation and evaluation of their classes (42% mean weight) and

evaluation of their compliance with administrative regulations (10% mean weight). Student ratings were given a mean weight of 9% with 42% assigning them a zero weight. Years of teaching experience (5% mean weight) and other types or sources of information tended to be weighted slightly--if weighted at all.

## Conclusions

The surveyed teachers were different from professional evaluators (Cochran-Smith & Lytle, 1990; Griswold, 1988; Gronlund, 1985; Gullickson & Ellwein, 1985; Hopkins & Stanley, 1990; Popham, 1990; and Stiggins, Frisbie & Griswold, 1989) in their approach to the evaluation of their students. These teachers appeared to: (1) prefer informal techniques to more formal techniques; (2) employ their observations and subjective analyses of homework, seatwork, and responses to oral questions for an undifferentiated combination of formative and summative purposes; (3) prefer evaluation techniques that required students to supply answers or produce complex behaviors--rather than select answers to multiple-choice or true-false/alternate response questions; and (4) believe that they should employ different evaluation and grading systems with students that differ. These teachers and many evaluators appeared to share the beliefs that: (1) grades should be based on student learning; and (2) tests are useful. The teachers, however, seemed to measure learning somewhat differently (and more informally)--while also wishing to establish

somewhat different standards for achievement for students who differ in perceived potential.

Questions?

Are these experienced teachers necessarily in error and in need of additional courses and workshops in evaluation and testing methodology? Marso and Pigge (1988a, 1988b, personal communications) have reported that teacher-produced test questions contain large numbers of technical errors (not too surprising to anyone who has taught testing classes or to anyone who constructs tests on a weekly or biweekly basis). However, teachers who were most confident concerning their evaluative skills--and who were rated most able by their supervisors--tended to produce test questions with the most errors. One explanation for such an apparent contradiction may lie in the possibility that experienced teachers, like experienced people in any field, have developed a more clinical approach to evaluation that may offset errors in information obtained from defective and inappropriate instruments. Hoge and Coladarci (1989) have reviewed the literature concerning the concurrent validity of teacher judgments of student achievement. Their analysis of 16 studies in which teacher judgments were compared with results from standardized achievement or ability tests indicated "generally high levels of agreement" (p. 308). The major moderator variable appeared to be the individual abilities of different teachers to make accurate

judgments--not the type of achievement measured or the characteristics of the students being judged. If teachers do need more instruction or coaching in the evaluation of student achievement, how much of this effort should be directed toward improvement in their skills as creators of formal tests--especially those that include the objectively-scored, selection questions that dominate the research literature? Would it be more appropriate to attempt to "coach" teachers to assist them to improve their frequently used, informal or subjective evaluative techniques that involve observation, oral questioning, analysis of seatwork and homework, and use of supply questions, essays, themes, and projects. Even the writing of essays might increase if teachers could be taught, somehow, to process the results in a more immediate manner that did not make major demands upon teachers' time.

Obvious Limitations

The teachers who were kind enough to respond to a clearly long and imperfect questionnaire were not necessarily typical teachers, and their often hurried responses probably contain most of the errors associated with self-reports. For example, why did only 67% report employing observation of student effort and motivation on a daily basis when it is hard to imagine any teacher not observing indicators of effort or motivation on a second-by-second or minute-by-minute basis (Kounin, 1970; Nicholl,

1979). The more real time and in-class analyses reported by Stiggins, Frisbie and Griswold (1989) should avoid many of the errors associated with self-report and survey approaches. However, many of the outcomes of their study are similar to those found by this survey. Teachers appear to prefer to employ measurement techniques that produce large and immediate quantities of information that may be of very questionable validity. The frequent use of homework as an indicator of achievement is particularly problematic since so many non-achievement factors may be associated with its completion. Similar validity problems may be associated with oral questions. Immediacy, subjectivity and richness of information appear to be preferred to the more delayed and more difficult to interpret (especially without item-analysis skills and support) information produced by selection questions.

Several of the more obvious questions that might be asked could include the following: (1) How valid are teacher judgments based upon the kinds of approaches that appear to be used most frequently by teachers? (2) What can be modified to increase the validity of these teacher judgments? Hoge and Caladarci's (1989) review and their references to studies of teachers' processing of information are steps in this direction. (3) If teachers would reduce the use of homework for summative judgements, how can teachers be assisted to improve the validity, range (beyond rote levels) and <u>efficient</u> use of more formed testing procedures?

Leiter (1976) has suggested that teachers already employ "objective" but "truncated" test score information to make subjective and context-based judgements--that are enriched and made more valid by the teacher's tacit knowledge. (4) If teachers can be assisted to employ more formal testing procedures in an effective manner, and if this addition does improve the validity of their judgements, how can they also be assisted to employ the information resulting from their procedures to resolve their need to combine an achievement standard with a need to treat and evaluate different students differently?

In Summation: We seem to know approximately what kinds of assessment procedures teachers employ. The procedures tend to be less formal than the procedures suggested in many measurement texts. We do not know how teachers convert information drawn from these approaches into judgments--or how valid and useful these judgments really are.

Postscript. Teachers and measurement specialists may face very different demands. The teacher-as-decision-maker must monitor an immediate and complex flow of information and make changes very quickly. The immediacy of the situation and its complex nature require simple and instantly applicable rules of procedure. Grading and reporting are only one of the many demands for attention--and one that can be ignored or manipulated for motivational or convenience reasons until the end of a grading

period. The management of activities (and the classroom order and the social and academic learning produced by these activities) is the ever-present pressure and the dominant obligation of the teacher. The measurement specialist is concerned with the valid and convenient (e.g., easily scored) measurement of large groups of strangers who may wait months for the opportunity to be presented with information describing their comparative position on an important but narrowly defined and very limited set of human characteristics. As measurement specialists move toward increasing utility, teachers should move toward increasing validity. The impact of any of these possible changes in procedures remains largely unknown.

30/11

References and Bibliography

Berliner, D. C. (1982). Instructional variables. In D. E. Orlosky (Ed.), Introduction to education. Columbus, OH: Merrill.

Biehler, R. F., & Snowman, J. (1986). Psychology applied to teaching. Boston: Houghton Mifflin.

Caffyn, R. E. (1989). Attitudes of British secondary school teachers and pupils to rewards and punishments. Educational Research, 31(3), 210-220.

Cochran-Smith, M., & Lytle, S. L. (1990). Research on teaching and teacher research: The issues that divide. Educational Researcher, 19(2), 2-17.

Cross, K. P., & Angelo, T. A. (1988). Classroom assessment techniques: A handbook for faculty (Tech. Rep. No. 88-A-004.0). Ann Arbor: University of Michigan (for the National Center for Research to Improve Postsecondary Teaching and Learning).

Fleming, M., & Chambers, B. (1983). Teacher-made tests: Windows on the classroom. New Directions for Testing and Measurement, 19, 29-38. San Francisco: Jossey-Bass.

Gage, N. L., & Berliner, D. C. (1988). Educational psychology (4th ed.). Boston: Houghton Mifflin.

Goldman, R. D., & Slaughter, R. E. (1976). Why college grade point average is difficult to predict. American Educational Research Journal, 11, 343-357.

Green, K. E., & Stager, S. F. (1986-1987). Testing: Coursework, attitudes and practices. Educational Research Quarterly, 11(2), 48-55.

Griswold, P. A. (1988). Teachers' attitudes toward testing: Recommendations for district policy and staff development. Portland, OR: Parkrose Public Schools, Office of Management Services. (ERIC Document Reproduction Service No. ED 305 363).

Gronlund, N. E. (1985). Measurement and evaluation in teaching (5th ed.). New York: Macmillan.

Gronlund, N. E., & Linn, J. E. (1990). Measurement and evaluation in teaching (6th ed.). New York: Macmillan.

Gullickson, A. R., & Ellwein, M. C. (1985). Post-hoc analysis of teacher-made tests: The goodness-of-fit between prescription and practice. Educational Measurement: Issues and Practice, 4, 15-18.

Gullickson, A. R. (1982). The practice of testing in elementary and secondary schools. Vermillion, SD: University of South Dakota. (ERIC Document Reproduction Service No. ED 229 391).

Gullickson, D. A. (1984). Teacher perspectives of their instructional use of tests. Journal of Educational Research, 77, 244-248.

Haladyna, T. M., Downing, S. M. (1989a). A taxonomy of multiple-choice item-writing rules. Applied Measurement in Education, 2(1), 37-49.

Haladyna, T. M., Downing, S. M. (1989b). Validity of a taxonomy of multiple-choice item-writing rules. Applied Measurement in Education, 2(1), 51-78.

Hall, B. W., Carroll, D., & Comer, C. B. (1988). Test use among classroom teachers and its relationship to teaching levels and teaching practices. Applied Measurement in Education, 1(2), 145-156.

Hoge, R. D., & Coladarei, T. (1989). Teacher-based judgments of academic achievement: A review of literature. Review of Educational Research, 59(3), 297-313.

Hopkins, K. D., Stanley, J. C., & Hopkins, B. R. (1990. Educational and psychological measurement and evaluation (7th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Kirby, P. C., & Oescher, J. (1989). Assessing teacher-made math and science tests. American Secondary Education, 18(1), 6-12.

Kounin, J. S. (1970). Discipline and group management in classrooms. New York: Holt, Rinehart, & Winston.

Kunder, L. H., & Porwoll, P. J. (1977). Reporting pupil progress: Policies, procedures, and systems. Arlington, VA: Educational Research Service.

Leiter, K. C. W. (1976). Teachers' use of background knowledge to interpret test scores. Sociology of Education, 49, 59-65.

Marso, R. N., & Pigge, F. L. (1988). An analysis of teacher-made tests: Testing practices, cognitive demands, and item construction errors. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans. (ERIC Document Reproduction Service No. ED 298 174).

Marso, R. N., & Pigge, F. L. (1988). Ohio secondary teachers' testing needs and proficiencies: Assessments by teachers, supervisors, and principals. American Secondary Education, 17(2), 2-9.

Marso, R. N., & Pigge, F. L. (1987). Teacher-made tests and testing: Classroom resources, guidelines, and practices. Paper presented at the annual meeting of the Midwestern Educational Research Association, Chicago. (ERIC Document Reproduction Service No. ED 291 781).

Milton, O., Pollio, H. R., & Eison, J. A. (1986). Making sense of college grades: Why the grading system does not work and what can be done about it. San Francisco: Jossey-Bass.

Nicholls, J. G. (1979). Quality and equality in intellectual development: The role of motivation in education. American Psychologist, 34, 1071-1084.

Popham, W. J. (1990). Modern educational measurement: A practitioner's perspective (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Stiggins, R. J., Frisbie, D. A., & Griswold, P. A. (1989). Inside high school grading practices: Building a research agenda. Educational Measurement: Issues and Practice, 8(2), 5-14.

Stiggins, R. J., Griswold, M. M., & Wikelund, K. R. (1989). Measuring thinking skills through classroom assessment. Journal of Educational Measurement, 26(3), 233-246.

Terwilliger, J. S. (1989). Classroom standard setting and grading practices. Educational Measurement: Issues and Practice, 8(2), 15-19.

Toledo Public Schools, & Toledo Federation of Teachers. (1985). The Toledo Plan: Intern intervention evaluation. Toledo, OH: Toledo Public Schools and the Toledo Federation of Teachers.

Woolfolk, A. E. (1990). Educational psychology (4th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Wood, P. H. (1980). The effects of test threat and test anxiety upon the continued academic interests of college students. Unpublished doctoral dissertation, Teachers College, Columbia University.

Wood, P. H., Stang, G., & Wood, J. (1980). Student ratings of college courses and college teachers: Reliability, generalizability, validity and utility. Bowling Green, OH: Bowling Green State University, College of Education. (ERIC Document Reproduction Service No. ED 193 263).

Yeh, J. P. (1980). A re-analysis of test-use data. Los Angeles, CA: Center for the Study of Evaluation, University of California, Los Angeles. (ERIC Document Reproduction Service No. ED 205 590).

30/11

Table 1

Descriptors of Teachers Who Responded to the Survey (n = 258)

| Experience | Mean | SD | Median | Range |
|---|---|---|---|---|
| Years taught | 12.8 | 8.0 | 12 | 1 (5%) to 34 |
| Number students assigned | 76 | 98 | 30 | 2 to 801 |
| Number handicapped students | 4.9 | 8.8 | 0 | 0 (35%) to 28 |
| Number special education classes taken | 2.3 | 3.4 | 0.5 | 0 (50%) to 9 (or more) |
| Number testing/evaluation classes taken | 1.6 | 1.1 | 1 | 0 (12%) to 6 |

| Grade Level Taught[1] | n | % |
|---|---|---|
| K-1 | 23 | 9 |
| 2 | 16 | 6 |
| 3 | 25 | 10 |
| 4 | 31 | 12 |
| 5 | 16 | 6 |
| 6 | 19 | 7 |
| 7 | 11 | 4 |
| 8 | 27 | 10 |
| 9 | 25 | 10 |
| 10-12 | 65 | 26 |

| Type of Class Taught | n | % |
|---|---|---|
| Regular-Academic | 199 | 77 |
| Special Education | 30 | 12 |
| Vocational | 15 | 6 |
| Performance (Music, Art, Physical Education) | 14 | 5 |

NOTE: [1]Many teachers teach students from several grade levels. Teachers were asked to focus upon a particular level or upon a particular course which could be categorized as most often falling within one grade level.

Table 2

Percentages of Students Assigned by Teachers to Letter Grade Categories

A. All Teachers (n = 242) — % of Students Assigned to Each Grade Category

| Grade Categories | Mean (SD) | Median | Modes (# Tchrs. at Modes) | Range |
|---|---|---|---|---|
| A grades | 25% (19%) | 20% | 10% (n = 53) | 0% to 90% |
| B grades | 31% (14%) | 30% | 20% & 30% (n = 44, 41) | 3% to 90% |
| C grades | 29% (15%) | 30% | 30% & 40% (n = 35, 31) | 0% to 90% |
| D grades | 10% (8%) | 10% | 10% (n = 51) | 0% to 40% |
| F grades | 5% (6%) | 5% | 0% & 5% (n = 68, 57) | 0% to 55% |
| A or B grades | 56% (22%) | 55% | 40% (n = 25) | 4% to 100% |
| D or F grades | 15% (13%) | 13% | 10, 0, 5 (n = 32, 29, 28) | 0% to 85% |

B. Sub-Groups of Teachers (by Type)

| Grade Categories | Reg. Tchrs. (n = 191) Mean (SD) | Spec. Educ. (n = 23) Mean (SD) | Vocational (n = 15) Mean (SD) | Performance (n = 13) Mean (SD) | F | p | r | Eta |
|---|---|---|---|---|---|---|---|---|
| A grades | 22% (17%) | 26% (24%) | 37% (23% | 42% (21%) | .72 | .0001 | .28 | .29 |
| B grades | 31% (14%) | 30% (15%) | 28% (11%) | 28% (11%) | .40 | .74 | -.06 | .07 |
| C grades | 30% (14%) | 33% (19%) | 22% (14%) | 18% (15%) | 4.0 | .008 | -.17 | .22 |
| D grades | 11% (8%) | 7% (10%) | 8% (5%) | 7% (7%) | 1.9 | .12 | -.14 | .15 |
| F grades | 6% (7%) | 3% (5%) | 4% (4%) | 4% (6%) | 1.5 | .23 | -.10 | .13 |
| A or B grades | 54% (22%) | 56% (25%) | 65% (21%) | 70% (25%) | 3.4 | .02 | .20 | .20 |
| D or F grades | 17% (13%) | 11% (14%) | 12% (8%) | 11% (13%) | 2.0 | .11 | -.14 | .16 |

(table continues)

Table 2 (continued)

C. Sub-Groups of Teachers (by Grade Level)

| Grade Categories | K-3 (n = 59) Mean (SD) | 4-6 (n = 62) Mean (SD) | 7-9 (n = 59) Mean (SD) | 10-12 (n = 63) Mean (SD) | F | p | r | Eta |
|---|---|---|---|---|---|---|---|---|
| A grades | 33% (19%) | 22% (18%) | 19% (16%) | 25% (20%) | 6.3 | .0004 | -.14 | .27 |
| B grades | 37% (13%) | 32% (14%) | 27% (14%) | 28% (12%) | 6.5 | .0003 | -.24 | .27 |
| C grades | 22% (13%) | 33% (17%) | 31% (14%) | 29% (14%) | 6.1 | .0005 | .13 | .27 |
| D grades | 6% (4%) | 9% (6%) | 15% (10%) | 11% (7%) | 17.6 | .0001 | .32 | .43 |
| F grades | 3% (3%) | 4% (5%) | 9% (9%) | 6% (5%) | 12.0 | .0001 | .25 | .35 |
| A or B grades | 69% (18%) | 54% (22%) | 46% (22%) | 54% (21%) | 13.0 | .0001 | -.27 | .37 |
| D or F grades | 8% (7%) | 12% (10%) | 24% (18%) | 17% (11%) | 18.0 | .0001 | .31 | .43 |

Table 3

Percentage Weights Assigned to Grade-Related Factors

| Factors | (n = 258) All Tchrs. Mean (SD) | (n = 258) Elem. Mean (SD) | (n = 102) Sec. Mean (SD) | (n = 164) Acad. Mean (DS) | (n = 18) Special Mean (SD) | (n = 9) Voc. Mean (SD) | (n = 11) Perf. Mean (SD) | F (El-Sec) | F (Type) |
|---|---|---|---|---|---|---|---|---|---|
| | Mean % Weights (and SD) Assigned to Factors by: | | | | | | | | |
| Achievement of Academic Objectives | 62% (26) | 65% (28) | 59% (24) | 67% (24) | 54% (32) | 50% (22) | 23% (16) | F = 3 | F = 13 |
| Motivation and Effort | 5% (7) | 6% (9) | 4% (6) | 6% (6) | 7% (11) | 5% (7) | 3% (7) | F = 5 | F = 1.7 |
| Attitudes | 2% (4) | 3% (5) | 2% (4) | 2% (3) | 1% (3) | 5% (6) | 12% (7) | F = 1.5 | F = 33 |
| Attendance | 2% (4) | 1% (2) | 2% (4) | 1% (3) | 2% (3) | 7% (5) | 5% (6) | F = 8 | F = 13 |
| Homework and Classwork | 17% (17) | 12% (14) | 22% (19) | 17% (18) | 16% (14) | 21% (15) | 14% (12) | F = 19 | F = .2 |
| Extra Credit | 2% (4) | 2% (3) | 2% (4) | 2% (4) | 2% (4) | 3% (4) | 2% (3) | F = .7 | F = .5 |
| Improvement | 5% (8) | 6% (9) | 3% (7) | 4% (7) | 11% (14) | 4% (5) | 9% (9) | F = 5 | F = 6 |
| Participation/Behavior in Class | 5% (7) | 5%(7) | 5% (8) | 4% (6) | 6% (8) | 6% (5) | 18% (15) | F = .2 | F = 18 |

NOTE: All values above 1 are rounded to the nearest integer.

Table 4

Evaluation Techniques and their Reported Frequency of Use

| Techniques | Reported Frequency of Use (% of Teachers: n = 200) | | | | | |
|---|---|---|---|---|---|---|
| | Daily = 5 | Weekly = 4 | Monthly = 3 | Sometimes = 4 | Not = 5 | Mean (SD) |
| Oral Questions | 60% | 8% | 4% | 10% | 17% | 3.8 (1.6) |
| Fill-in/Completion Questions | 14% | 43% | 17% | 16% | 9% | 3.4 (1.2) |
| Matching Questions | 5% | 32% | 27% | 22% | 14% | 2.9 (1.1) |
| True-False Questions | 2% | 18% | 29% | 21% | 29% | 2.4 (1.2) |
| Multiple-Choice Questions | 5% | 32% | 35% | 16% | 12% | 3.0 (1.1) |
| Problem Questions | 19% | 31% | 13% | 21% | 14% | 3.2 (1.4) |
| Short Essay Questions | 3% | 26% | 15% | 23% | 34% | 2.4 (1.3) |
| Longer Essay Questions | 1% | 3% | 10% | 15% | 72% | 1.5 (0.8) |
| Papers-Reports-Themes | 3% | 4% | 23% | 35% | 36% | 2.0 (1.0) |
| Projects | 6% | 10% | 59% | 34% | 19% | 2.5 (1.1) |
| Performances | 4% | 9% | 18% | 27% | 42% | 2.1 (1.2) |
| Homework Assignments | 45% | 33% | 5% | 11% | 6% | 4.0 (1.2) |
| In-class/Seatwork | 63% | 25% | 1% | 7% | 3% | 4.4 (1.0) |
| Observation of Effort/Motivation | 67% | 13% | 2% | 10% | 8% | 4.2 (1.4) |
| Sociograms | 0% | 1% | 3% | 26% | 70% | 1.4 (0.6) |
| Anecdotal Record | 6% | 11% | 10% | 27% | 47% | 2.0 (1.2) |
| Interest Inventories | 0% | 3% | 6% | 46% | 45% | 1.7 (0.7) |
| "Guess-Who" Questions | 1% | 3% | 6% | 25% | 65% | 1.5 (0.8) |
| Self-Esteem Measures | 6% | 15% | 6% | 30% | 43% | 2.1 (1.3) |
| Attitude Questionnaires | 0% | 4% | 4% | 39% | 54% | 1.6 (0.7) |

Table 5

Teacher Agreement with 17 Evaluation-Related Statements

| Statements | 1 = SA | 2 = A | Agree (1 or 2) | 3 = D | 4 = SD | Disagree (3 or 4) | 5 = ? |
|---|---|---|---|---|---|---|---|
| 1. Grades on a report card (for the subjects I teach most) should reflect how much students have learned. | 33% | 59% | 93% | 3% | 1% | 4% | 2% |
| 2. Teachers should set different objectives for different students and grade according to how well these different objectives are met. | 27% | 38% | 65% | 23% | 5% | 28% | 7% |
| 3. I tend to grade my less able students more leniently. | 9% | 46% | 55% | 32% | 11% | 43% | 2% |
| 4. Attendance should be included in subject grades. | 10% | 36% | 46% | 33% | 16% | 49% | 5% |
| 5. I would include more projects in my class if I had more time for grading. | 23% | 47% | 70% | 18% | 4% | 22% | 8% |
| 6. District-wide testing tends to improve education. | 0% | 10% | 10% | 37% | 30% | 67% | 23% |
| 7. Grades on report cards should reflect effort as well as levels of achievement. | 34% | 51% | 85% | 12% | 1% | 13% | 1% |
| 8. If I had more time for grading, I would have students do more writing on examinations. | 21% | 41% | 62% | 24% | 6% | 30% | 8% |
| 9. I am uncomfortable with my grading procedures. | 3% | 10% | 14% | 48% | 38% | 86% | 0% |

(table continues)

Table 5 (continued)

| | Statements | 1 = SA | 2 = A | Agree (1 or 2) | 3 = D | 4 = SD | Disagree (3 or 4) | 5 = ? |
|---|---|---|---|---|---|---|---|---|
| 10. | I tend to set different obejctives for students of differing abilities. | 8% | 43% | 61% | 30% | 8% | 38% | 1% |
| 11. | Report card grades tend to be more accurate when they are based mostly on test or quiz scores. | 4% | 33% | 37% | 40% | 15% | 55% | 7% |
| 12. | Tests tend to do more harm than good. | 2% | 9% | 12% | 69% | 13% | 82% | 6% |
| 13. | State-wide testing tends to improve education. | 0% | 8% | 8% | 36% | 32% | 68% | 24% |
| 14. | I tend to have different assignments for students of different ability levels. | 16% | 36% | 52% | 38% | 9% | 47% | 2% |
| 15. | Completion of homework should be included in subject grades. | 39% | 52% | 91% | 4% | 2% | 6% | 2% |
| 16. | If possible, I would not give any tests at all. | 6% | 13% | 19% | 53% | 25% | 78% | 2% |
| 17. | The teachers who teach similar classes (in my school) tend to use similar grading systems. | 13% | 56% | 68% | 7% | 1% | 8% | 23% |

Table 6

Teachers' Preferred Weighting (%) of Methods Used to Evaluate Their Teaching

| | Percent Weighting Preferred (n = 186) | | | |
|---|---|---|---|---|
| Method/Source | Mean (SD) | Median | Range | % at 0% Weight |
| Principal Visits | 42% (26%) | 40% | 0% to 100% | (4%) |
| Teacher Visits | 13% (15%) | 10% | 0% to 50% | (39%) |
| Student Ratings | 9% (12%) | 5% | 0% to 70% | (42%) |
| Parents' Ratings | 4% (6%) | 0% | 0% to 30% | (60%) |
| Students' Standard Test Scores | 5% (8%) | 0% | 0% to 50% | (63%) |
| Extra-Curricular Activities | 2% (6%) | 0% | 0% to 40% | (74%) |
| Compliance Adm. Regs. | 10% (10%) | 10% | 0% to 50% | (20%) |
| Service: School Committees | 4% (6%) | 0% | 0% to 25% | (56%) |
| Years of Teaching Experience | 5% (8%) | 0% | 0% to 50% | (59%) |
| Courses/CEUs Completed | 5% (8%) | 0% | 0% to 50% | (57%) |

## A Survey of Teachers

Surprisingly little is known about how teachers test and grade their students. This survey is intended to map some of this unknown area by asking you to tell us some of your opinions and procedures. Thank you for your time and help.

Background Information. People with different types of experience may have different views. Please answer the following questions by writing the NUMBER of the best answer on the short line in front of each question number.

_____ 1. Sex/Gender? 1 = Female 2 = Male
_____ 2. Teaching Status? 1 = Presently teaching
2 = Have taught previously
_____ 3. Approximately how many years have you taught? 1 = 1, 2 = 2, etc.
_____ 4. Approximately how many handicapped (special) students have you taught or tutored in the past year (or most recent year of teaching)? 0 = 0, 1 = 1, etc.
_____ 5. Approximately how many testing or measurement classes have you taken at a university? 0 = 0, 1 = 1, 2 = 2, ... 5 = 5, etc.
_____ 6. Approximately how many special education classes have you taken at a university? 0 = 0, 1 = 1, etc.
_____ 7. Grade Level? In what grade level do you now teach or have had the most recent experience? 99 = K, 1 = first grade, 2 = second grade, etc... 10 = soph., 12 = senior (STUDENTS: PICK THE MOST LIKELY GRADE LEVEL)
_____ 8. Approximately how many students do you teach this term (or the last term you taught)?
_____ 9. Write the names of the subjects which you teach now or have most recently taught. Elementary teachers write ELEMENTARY; secondary teachers write English, history, art, band, special education, etc.
(STUDENTS: Pick the MOST LIKELY SUBJECT(S) )

LETTER GRADES

For this section on LETTER GRADES, please identify a single subject (e.g., reading, English, mathematics, algebra, social studies, American history) and a single grade level (e.g., third grade, seventh grade, sophomore).

Subject _____ Grade level _____

Answer the following questions as they apply to the single subject and grade level you wrote above.

If your school system required you to give a letter grade for the subject you listed above, approximately what percentage of students in a typical class would you expect to receive the following grades?

_____% would probably get "A" grades
_____% would probably get "B" grades
_____% would probably get "C" grades
_____% would probably get "D" grades
_____% would probably get "F" or failing grades

Note: The percentages above should sum to 100%.

Assume you had to give your students grades of A, B, C, D, or F at the end of a grading period. In the single SUBJECT and grade level you listed, approximately what percentage of the final grade would come from each of the following areas?

_____% The student's demonstrated achievement of the course's academic objectives (test scores, project grades, daily work, etc.).
_____% The student's motivation to learn and effort to learn.
_____% The student's attitudes toward the class and toward school.
_____% The student's attendance (absences, tardiness).
_____% The student's completion of assigned homework and classwork.
_____% The student's completion of "extra credit" assignments.
_____% The student's improvement or working up to potential.
_____% The student's participation/behavior in class.

EVALUATION TECHNIQUES

Some of the following testing and measurement procedures are used by most teachers. Indicate which you would be likely to use daily, weekly, monthly, or not at all in the subject and grade level already mentioned.

1 = Daily: would use nearly every day
2 = Weekly: would use once or twice in most weeks
3 = Monthly: would use once or twice in most months
4 = Sometimes: would use--but not very often
5 = Not use: would probably not use for this class and level

Use the above numbers to indicate your likely frequency of use of:

Tests, Quizzes, Exams

_____ 1. oral questions
_____ 2. fill-in or completion
_____ 3. matching questions
_____ 4. T-F questions
_____ 5. multiple-choice questions
_____ 6. problem questions
_____ 7. short essay questions
_____ 8. longer essay questions

Personal-Social Measures

_____15. Sociograms
_____16. Anecdotal records
_____17. Interest inventories
_____18. "Guess-who" questions
_____19. Self-esteem measures
_____20. Attitude questionnaires
_____21. Other _____

Other

_____ 9. papers (report, themes, term papers)
_____10. projects (products: woodwork, maps, artwork)
_____11. performances (athletic, singing, speeches, presentations, skits)
_____12. homework assignments
_____13. in-class work or "seat work"
_____14. observations of student effort and motivation

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

_____ A. Which of the above approaches tends to give you the best idea of how well objectives in your class have been met at the end of each grading period? (Write the numbers of the one, two or three best techniques in the space to the left of the A).
_____ B. Which of the above is best for telling you, the teacher, if students are learning?
_____ C. Which would you use more if you knew more about it?
_____ D. Which would you use more if you had more time?
_____ E. Which procedure(s) seem most useful in helping your students to learn?
_____ F. Which procedure(s) are most helpful when you have to make out report cards?

Opinions. Indicate your agreement--or disagreement--with the following statements about testing and grading.

1 = SA = Strongly Agree
2 = A = Agree
3 = D = Disagree
4 = SD = Strongly Disagree
5 = ? = I really can't answer this question

_____ 1. Grades on a report card (for the subjects I teach most) should reflect how much students have learned.

_____ 2. Teachers should set different objectives for different students and grade according to how well these different objectives are met.

_____ 3. I tend to grade my less able students more leniently.

_____ 4. Attendance should be included in subject grades.

_____ 5. I would include more projects in my class if I had more time for grading.

_____ 6. District-wide testing tends to improve education.

_____ 7. Grades on report cards should reflect effort as well as levels of achievement.

_____ 8. If I had more time for grading, I would have students do more writing on examinations.

_____ 9. I am uncomfortable with my grading procedures.

_____10. I tend to set different objectives for students of differing abilities.

_____11. Report card grades tend to be more accurate when they are based mostly on test or quiz scores.

_____12. Tests tend to do more harm than good.

_____13. State-wide testing tends to improve education.

_____14. I tend to have different assignments for students of different ability levels.

_____15. Completion of homework should be included in subject grades.

_____16. If possible, I would not give any tests at all.

_____17. The teachers who teach similar classes (in my school) tend to use similar grading systems.

* * * * * * * * * * *